

Initiation au logiciel SAS

M1 MAS - Année universitaire 2023-2024

Alexie Robert

22 septembre 2023

Les objectifs du TP d'aujourd'hui

- Représenter des graphiques avec les procédures GCHART, GPLOT et BOXPLOT ;
- enregistrer les sorties avec l'Output Delivery System ;
- utiliser les procédures MEANS, UNIVARIATE, CORR et FREQ ;
- exporter vers des fichiers Excel ;
- découvrir l'instruction ARRAY et les boucles;

Les procédures graphiques et l'export vers un fichier pdf

Les procédures graphiques et l'export vers un fichier pdf

1 le programme

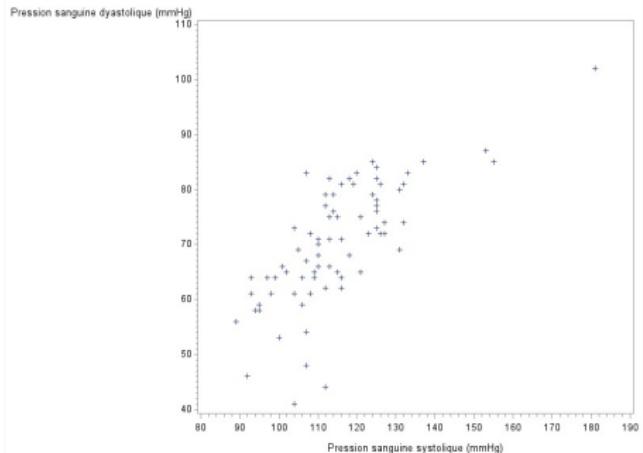
```
/* Graphique au format pdf */
ODS PDF FILE = "&path\graph.pdf";
PROC GPLOT DATA=TP.herp;
PLOT DIA*SYS;
RUN;
ODS PDF CLOSE;
```

Les procédures graphiques et l'export vers un fichier pdf

1 le programme

```
/* Graphique au format pdf */
ODS PDF FILE = "&path\graph.pdf";
PROC GPLOT DATA=TP.herp;
PLOT DIA*SYS;
RUN;
ODS PDF CLOSE;
```

2 la sortie



La procédure MEANS

1 le programme

```
PROC MEANS DATA=tp.herp;  
  VAR AGE ;  
  RUN;
```

La procédure MEANS

1 le programme

```
PROC MEANS DATA=tp.herp;  
  VAR AGE ;  
  RUN;
```

2 la sortie

Variable d'analyse : AGE Age de l'individu				
N	Moyenne	Ec-type	Minimum	Maximum
80	34.3500000	13.1756392	12.0000000	73.0000000

La procédure MEANS

1 le programme

```
□ PROC MEANS DATA=tp.herp SUM MEDIAN;  
    VAR AGE ;  
    RUN;
```

La procédure MEANS

1 le programme

```
□ PROC MEANS DATA=tp.herp SUM MEDIAN;
  VAR AGE ;
  RUN;
```

2 la sortie

Variable d'analyse : AGE Age de l'individu	
Somme	Médiane
2748.00	32.0000000

La procédure CORR

1 le programme

```
PROC CORR DATA=tp.herp;  
  VAR AGE SYS ;  
  RUN;
```

La procédure CORR

1 le programme

```
PROC CORR DATA=tp.herp;
  VAR AGE SYS ;
  RUN;
```

2 la sortie

La procédure CORR

2 Variables : AGE SYS

Statistiques simples							
Variable	N	Moyenne	Ec-type	Somme	Minimum	Maximum	Libellé
AGE	80	34.35000	13.17564	2748	12.00000	73.00000	Age de l'individu
SYS	80	114.85000	14.67116	9188	89.00000	181.00000	Pression sanguine systolique (mmHg)

Coefficients de corrélation de Pearson, N = 80
Proba > |r| sous H0: Rho=0

		AGE	SYS
AGE	Age de l'individu	1.00000	0.39744 0.0003
SYS	Pression sanguine systolique (mmHg)	0.39744 0.0003	1.00000

La procédure FREQ

1 le programme

```
/* Effectifs et fréquences,  
avec les formats */  
PROC FREQ DATA=tp.herp;  
TABLE SEXE SURPOIDS;  
FORMAT SURPOIDS $SURP. SEXE $SEX.;  
RUN;
```

La procédure FREQ

1 le programme

```
/* Effectifs et fréquences,
avec les formats */
PROC FREQ DATA=tp.herp;
TABLE SEXE SURPOIDS;
FORMAT SURPOIDS $SURP. SEXE $SEX. ;
RUN;
```

2 la sortie

Sexe de l'individu				
SEXE	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
Garçon	40	50.00	40	50.00
Fille	40	50.00	80	100.00

Surpoids de l'individu				
SURPOIDS	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
Surpoids	43	53.75	43	53.75
Absence de surpoids	37	46.25	80	100.00

La procédure FREQ

1 le programme

```
/* Test du chi-deux */
PROC FREQ DATA=tp.herp;
TABLE SEXE*SURPOIDS / NOPERCENT NOCOL CHISQ;
FORMAT SURPOIDS $SURP. SEXE $SEX.;
RUN;
```

La procédure FREQ

1 le programme

```
/* Test du chi-deux */
PROC FREQ DATA=tp.herp;
TABLE SEXE*SURPOIDS / NOPERCENT NOCOL CHISQ;
FORMAT SURPOIDS $SURP. SEXE $SEX.;
RUN;
```

2 la sortie

Fréquence Pct de ligne	Table of SEXE by SURPOIDS			
	SEXE(Sexe de l'individu)	SURPOIDS(Surpoids de l'individu)		
		Surpoids	Absence de surpoids	Total
	Garçon	24 60.00	16 40.00	40
	Fille	19 47.50	21 52.50	40
	Total	43	37	80

Statistiques pour la table de SEXE par SURPOIDS

Statistique	DDL	Valeur	Prob
Khi-2	1	1.2571	0.2622

L'instruction ARRAY

1 la table

Obs.	id	x
1	a	0001
2	b	1010
3	c	1010
4	d	0011
5	e	1100
6	f	

L'instruction ARRAY

1 la table

Obs.	id	x
1	a	0001
2	b	1010
3	c	1010
4	d	0011
5	e	1100
6	f	

2 le programme

```
/* substr(x,i,j) : extrait j caractères à partir du ième caractère
dans la colonne x */
/* x_1-x_4 équivalent de x_1 x_2 x_3 x_4 */
/* rf(1) = x_1 ... */

data a;
set a;
array rf (4) $1 x_1 x_2 x_3 x_4 ;
do i=1 to 4 ;
rf(i)=substr(x,i,1);
end ;
run;
```

L'instruction ARRAY

la sortie

Obs.	id	x	x_1	x_2	x_3	x_4	i
1	a	0001	0	0	0	1	5
2	b	1010	1	0	1	0	5
3	c	1010	1	0	1	0	5
4	d	0011	0	0	1	1	5
5	e	1100	1	1	0	0	5
6	f						5